

Basis Set Modeling for Molecular Calculations Using Effective Core Potential

MARCELO GIORDAN*, ROGÉRIO CUSTODIO

Instituto de Química, Universidade Estadual de Campinas, UNICAMP, C.P. 6154, 13081-970, Campinas São Paulo, Brazil

Received 17 February 1997; accepted 20 June 1997

ABSTRACT: A new approach for developing of basis sets to be used along with effective core potential is systematically studied. The behavior of the LCAO coefficients versus the $\ln(\alpha)$ of the respective primitives can provide simple guidelines to establish the range over which the basis set should be developed or modified, especially when using effective core potential. Double-zeta basis sets were modeled for SBK pseudopotential from all-electron basis sets for a series of compounds containing elements of the second period of the periodic table. Application of the modeled basis sets at the Hartree–Fock and MP2 levels of theory shows that the new method provides molecular properties as accurate as those calculated by all-electron calculations. © 1997 John Wiley & Sons, Inc. *J Comput Chem* 18: 1918–1929, 1997

Keywords: effective core potential; basis functions; generator coordinate method; molecular properties; MP2

Introduction

The search for adequate basis sets is still a problem of great interest for theoreticians or users of *ab initio* and density functional programs. Although a large number of basis sets is available in the literature, the interest for accurate, compact, and well-adapted basis functions is a subject that has been addressed in recent works.^{1–6}

* Permanent address: Departamento de Metodologia do Ensino e Educação Comparada, Faculdade de Educação, Universidade de São Paulo, Av. da Universidade, 308, CEP 05508-900, São Paulo SP, Brazil

Correspondence to: R. Custodio

Contract/grant sponsor: Fundação de Amparo à Pesquisa do Estado de São Paulo; contract/grant number: 1996/1102-4

Contract/grant sponsor: Conselho Nacional de Desenvolvimento Científico e Tecnológico

Contract/grant sponsor: Fundo de Apoio à Pesquisa

A general overview on the formulation of basis sets shows that little progress has been made in the development of new techniques. New basis sets have been developed mainly through the application of the traditional optimization of exponents in the atoms. The atom-optimized exponents are then transferred to molecules and enlarged with diffuse and/or polarization functions to improve the description of the electronic distribution in the molecular environment. This standard procedure can provide accurate molecular properties at the Hartree–Fock or at the correlated level of theory. The two most common problems to overcome in modeling a good basis set are: (a) the presence of local minima is often found when the exponents are optimized against the electronic energy and (b) when the energetic balance in the region near the nuclei tends to concentrate a large number of primitives in this region. Whereas (a) corresponds to a technical problem, the computational limitations imposed by (b) restrict the size of the molecules that can be studied. The usually large size of good basis sets tends to confine accurate *ab initio* calculations to small molecules. A simple and efficient way to reduce the basis set size is to replace the core electrons by an effective core potential. Pseudopotentials can also be used to include some relativistic effects when necessary.^{7–9}

Presently, two of the most common pseudopotentials used are the Hay and Wadt (HW)^{10,11} and the Stevens, Basch, and Krauss (SBK)^{12,13} effective pseudopotentials. These core potentials are followed by very compact basis sets obtained by the standard procedure just described. Calculations for some molecules and ions, either at the Hartree–Fock level or including correlation effects, fail to provide chemically reasonable properties. Very often the low quality of the results is more strongly associated with the performance of the basis set than with the pseudopotential itself. In this sense, better basis sets as well as polarization and diffuse functions should be developed.

A simple procedure based on the behavior of the linear combination of atomic orbital coefficients (LCAO) of uncontracted basis set has provided a simple way to detect the range over which the exponents of the basis set should be concentrated. This procedure is formally called the generator coordinate method (GCM)^{14,15} and has been successfully applied to model or develop basis sets in the atomic and molecular environment.^{16,25}

The objective of this article is to study systematically the applicability of GCM to model basis sets to be used along with the pseudopotential. The SBK compact effective pseudopotential^{12,13} is employed in this work. In the next section, the theory involving the effective core potential associated with the generator coordinate method is described. The third section outlines the method to model the basis sets. The method is then applied to calculate some properties (geometrical parameters, HOMO and LUMO energies, harmonic vibrational frequencies, IR intensities, and dipole moments) of molecules containing elements of the first row of the periodic table at the Hartree–Fock and MP2 levels of theory. The results are analyzed through some simple statistical criteria and compared with accurate calculations and experimental data. Special attention is given to the performance of the calculations carried out with the modified basis sets used along with the pseudopotential in comparison to all-electron calculations.

Theory

The generator coordinate method (GCM)^{14,15} suggests that the one-electron functions should be represented by integral transforms. For one-center systems, these functions are represented by the ansatz:

$$\psi_i(1) = \int_0^\infty f_i(\alpha) \phi_i(\alpha, \mathbf{r}_1) d\alpha \quad (1)$$

where $\phi_i(\alpha, 1)$ is the basis function with exponent or generator coordinate α and $f_i(\alpha)$ is a weight function. The application of eq. (1) to calculate expectation values using a Slater determinant leads to the Griffin–Hill–Wheeler–Hartree–Fock (GHWHF) equations of the form:

$$\int_0^\infty [F(\alpha, \beta) - \varepsilon_i \cdot S(\alpha, \beta)] f_i(\beta) d\beta = 0 \quad (2)$$

In eq. (2), S and F are, respectively, the overlap and Fock kernels.¹⁶

Eq. (2) can be solved numerically by an integral discretization (GCM-ID) or numerical integration technique. The ID technique transforms eq. (2) into a set of equations similar to the Hartree–Fock–Roothaan equation:

$$\sum_{l=1}^N [F(\alpha_k, \beta_l) - \varepsilon_i S(\alpha_k, \beta_l)] c_l(\beta_l) = 0 \quad (3)$$

Comparison of eq. (3) and eq. (2) shows that the weight functions can be identified with the LCAO coefficients of the Hartree–Fock–Roothaan equations. The precise correlation between a weight function and the respective set of LCAO coefficients will depend on the quadrature method employed. However, in practice, the most significant aspect is that the discretization of eq. (2) using Gaussians as generator functions can associate the LCAO coefficients as an approximation of the continuous weight functions, $f_i(\alpha)$, and the set of exponents as points necessary to integrate numerically the one-electron functions [Eq. (1)].

Eq. (1) shows that the complete space of the α exponents should comprise the range from 0 to ∞ . Although, in practice, the α space is limited to a narrower region, a large number of functions must be included in an all-electron calculation either in atomic or molecular environments to describe approximately the regions near the nuclei (the region of large exponents).

The change of the all-electron Fock operator to a valence effective Fock operator confines the electronic distribution to the valence pseudo-orbitals. The inner electrons are now represented by a pseudopotential that depends on the particular way the electrons were removed, and consequently, the appearance of the valence orbitals. The two main procedures to simulate the inner electrons are based on: (a) the Phillips–Kleinman pseudopotential method,²⁶ which inspired the SBK and HW pseudopotentials; and (b) the model potential proposed by Bonifacic and Huzinaga.²⁷ In the model potential method the valence pseudo-orbitals are described with as many nodes as the orbitals from all-electron calculations. On the other hand, the SBK and HW pseudopotentials describe the atomic pseudo-orbitals as nodeless functions. The main advantage with respect to the nodeless pseudo-orbitals is the smaller number of primitive functions required to describe the electronic distribution of the valence region, consequently reducing the computational time to evaluate one- and two-electron integrals.

In a conventional Hartree–Fock calculation one of the main conditions to consider in the use of the pseudopotential is the similarity of the electronic distribution between a molecular or atomic orbital calculated by an all-electron procedure with the equivalent pseudo-orbital obtained by some of the methods based on the pseudopotential. Considering the correspondence between the set of LCAO coefficients in an all-electron calculation with the weight functions from GCM, it is natural to ask if

the pseudo-orbitals can also be represented by an integral transform representation like eq. (1). The answer is not obvious because the weight functions are unknown and can only be obtained variationally. The requirement for a particular Hartree–Fock model to employ the integral transform is that the set of LCAO coefficients for the pseudo-orbitals be representations of a smooth and continuous function vanishing at the boundaries of the α space. Comparisons between all-electron and pseudo-orbital weight functions for some atoms and simple molecules using SBK pseudopotential^{23–25} showed an equivalent behavior of the electronic distributions in the α space. An interesting information obtained from these comparisons is a particular α exponent (α_0) characterized as the region where the pseudopotential starts to take effect. This α_0 can be used as an inner point at which one starts to model a basis set. In the light of GCM it can be said that eq. (1) should be integrated from 0 to an α_0 where the basis set has a minimal effect on the wave function because it is replaced by a pseudopotential. Mathematically, the use of a pseudopotential changes eq. (1) into:

$$\psi_i(1) = \int_0^{\alpha_0} f_i(\alpha) \phi_i(\alpha, \mathbf{r}_1) d\alpha \quad (4)$$

and consequently, eq. (2) has to be integrated in a more restricted space.

An important point must be analyzed before the tendency of the LCAO coefficients in a pseudo-orbital be considered as approximated weight functions. An important property to be observed by solution of the GHWHF equations is the requirement of square integrable weight functions. That is, the norm of f_i must be finite.²⁸ Eq. (2) can be rewritten in terms of discretized pseudo-orbitals represented on an orthonormal basis:

$$\sum_{m=1}^N F_{nm} C_m = E C_n \quad (5)$$

In such a representation the weight function is given by²⁸:

$$f(\alpha_k) = \sum_{n=1}^N \frac{C_n u_n(\alpha_k)}{\sqrt{\lambda_n}} \quad (6)$$

where u_n and λ_m are, respectively, eigenvectors and eigenvalues of the overlap matrix, and C_n the LCAO coefficient of a pseudo-orbital. The weight functions represented by eq. (6) will be square integrable depending on the relative diminution of

$|C_n|^2$ and λ_n . It is interesting to note that comparisons of all-electron with pseudopotential calculations using all-electron basis sets show a significant diminishment of the LCAO coefficients above α_0 when core potentials are used with respect to the all-electron results. If the all-electron calculations can be considered as discretized representations of the GHWHF equations, the more accentuated reduction of the LCAO coefficients in the core region indicates that GCM should not be used only in terms of the discretization of the valence region, but also to define a cutoff value for the exponent space in the core region.

In practice, the main problem is to detect α_0 and to model or modify a basis set in such a way that the desired molecular properties can be well described. The simplest way to achieve this is to perform all-electron calculations in atoms and compare the resulting weight functions with calculations using an effective core potential. Once α_0 has been detected, the basis set can be conveniently adapted to a molecular environment through the analysis of the LCAO coefficients (the approximated weight functions). The basic procedure is described below.

Basis Set Development

The use of the SBK effective core potential (ECP) in atoms represented by all-electron uncontracted Gaussian basis sets causes significant changes in the linear combination of atomic orbital coefficients (LCAO) associated with the primitives of large exponents (core region). The simplest way to visualize these changes is to plot the LCAO coefficients of the uncontracted basis set versus the logarithm of the respective exponents $[\ln(\alpha)]$ of the various s, p, d, ... symmetries of the basis functions. Figure 1 shows the plot of the LCAO coefficients for the 2s and 2p atomic orbitals versus $\ln(\alpha)$ for the $C(^3P)$ atom using two basis sets, (11s, 7p) and (5s, 6p).²⁹ The set of 2s and 2p LCAO coefficients (or the discretized approximation of weight functions) shows a continuous and smooth behavior with $\ln(\alpha)$, approaching to zero in the core region [large $\ln(\alpha)$]. This behavior indicates that the contribution of the primitives with exponents larger than the cutoff e^x (depicted by a dashed line in Fig. 1) is vanishingly small. On the other hand, a comparison of the weight function calculated at the Hartree–Fock level including all-electrons with the SBK core potential calculations

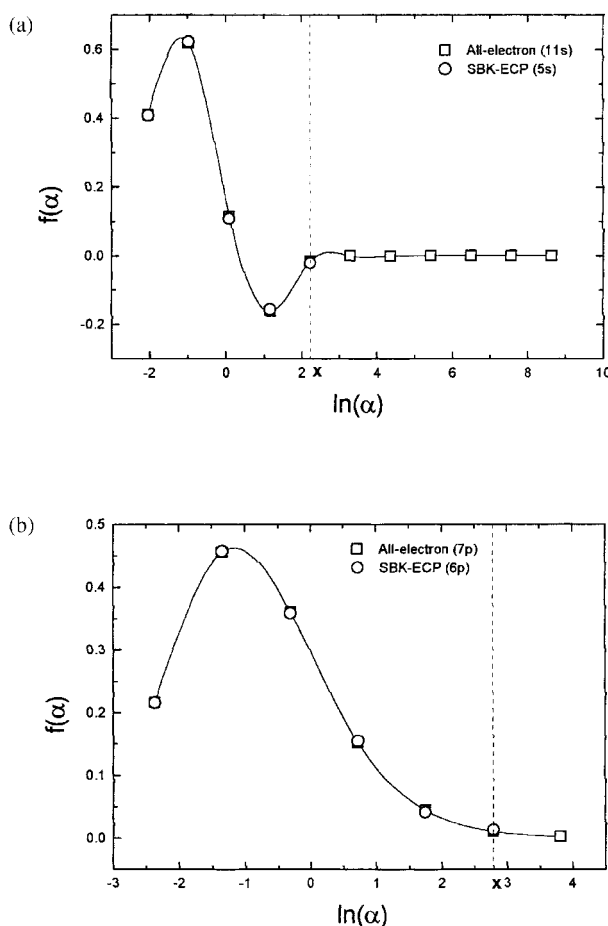


FIGURE 1. Weight functions for the 2s (a) and 2p (b) atomic orbitals of $C(^3P)$ calculated at the Hartree–Fock level with different basis sets.

shows a significant discrepancy between both functions for $\ln(\alpha) > x$. The approximated weight function calculated with SBK yields values close to zero for exponents larger than e^x , whereas all-electron weight function presents significant values. In this sense, x is clearly indicative of a region where the pseudopotential starts to act. For the 2p orbital, only one exponent is greater than e^x and the respective primitive presents a small contribution to the electronic energy. A convenient value of LCAO coefficients to be used as cutoff is 0.01. Therefore, the exponents of the weight functions with values smaller than 0.01 were eliminated from the original basis set (GCDZ), reducing its size from (11s, 7p) to (5s, 6p), hereafter called GCDZ-SBK. Plots of the weight functions for $B(^2P)$, $N(^4S)$, $O(^3P)$, and $F(^2P)$ exhibit a behavior similar to $C(^3P)$ (data not shown).

Table I shows the total energy and the valence orbital energies for $B(^2P)$, $C(^3P)$, $N(^4S)$, $O(^3P)$, and

TABLE I.
Total SCF Energy and Orbital Energies (a.u.) for
Atomic ECP-SBK Calculations.

Atom	Basis Sets	$-E_{\text{total}}$	$-E_{2s}$	$-E_{2p}$
B(³ P)	(11, 7)	2.53868	0.4946	0.3098
	(5, 6)	2.53809	0.4945	0.3096
	ref. 6		0.4947	0.3099
C(³ P)	(11, 7)	5.31575	0.7061	0.4331
	(5, 6)	5.31449	0.7058	0.4331
	ref. 6		0.7056	0.4334
N(⁴ S)	(11, 7)	9.64368	0.9457	0.5678
	(5, 6)	9.64150	0.9454	0.5678
	ref. 6		0.9453	0.5675
O(³ P)	(11, 7)	15.66460	1.2448	0.6320
	(5, 6)	15.66059	1.2445	0.6319
	ref. 6		1.2443	0.6319
F(² P)	(11, 7)	23.87545	1.5732	0.7296
	(5, 6)	23.86836	1.5726	0.7294
	ref. 6		1.5725	0.7300

F(²P) calculated at the Hartree–Fock level using the pseudopotential with the uncontracted GTO (11s, 7p) and modified (5s, 6p) basis sets. The valence orbital energies obtained with high quality STO basis sets for all-electron calculations are also presented.⁶ The difference between the valence orbital energies calculated with both GTO basis sets is negligible and in excellent agreement with the SCF all-electron STO basis set calculations.

The use of the (5s, 6p) basis sets in molecules requires a contraction scheme. Dunning’s segmented method³⁰ was adopted in this work. Double-zeta basis sets were generated from the uncontracted (5s, 6p) basis sets. The contraction coefficients were obtained from SCF-LCAO coefficients from atomic calculations. The (4, 1; 4, 2) scheme provided, in general, the smallest loss in the total energy of diatomic homonuclear molecules when compared with the same calculation using the uncontracted basis set. The exponents and contraction coefficients are shown in Table II.

It is well established that polarization functions are essential for a good description of the molecular and atomic properties. Different sets of d and f polarization functions were obtained for homonuclear diatomic molecules using the minimum energy criterion. The exponents were optimized at the Hartree–Fock level in the electronic ground

TABLE II.
Exponents and Contraction Coefficients for
GCDZ-SBK Basis Sets

Atom	α_{2s}	$C_{i, 2s}$	α_{2p}	$C_{i, 2p}$
B(² P)	6.664029	−0.01456	10.355243	0.01216
	2.289048	−0.14669	3.716597	0.03688
	0.786273	0.05232	1.333922	0.13713
	0.270079	0.60040	0.478758	0.33744
	0.092770	1.00000	0.171831	0.47265
			0.061672	0.23634
C(³ P)	9.301372	−0.02136	16.215977	0.01298
	3.204431	−0.15562	5.782145	0.04149
	1.103964	0.10748	2.061744	0.15454
	0.380328	0.62270	0.735158	0.35817
	0.131028	1.00000	0.262136	0.45730
			0.093470	0.21570
N(⁴ S)	12.132434	−0.02841	23.046581	0.01065
	4.180818	−0.16175	8.189444	0.03493
	1.440704	0.16187	2.910062	0.12934
	0.496464	0.63329	1.034071	0.28594
	0.171081	1.00000	0.367450	0.34298
			0.130571	0.15690
O(³ P)	17.094682	−0.02512	28.531470	0.01800
	5.908556	−0.175298	9.952746	0.06059
	2.042216	0.18096	3.471856	0.22166
	0.705865	0.60958	1.211101	0.43452
	0.243973	1.00000	0.422473	0.40025
			0.147373	0.11010
F(² P)	22.374077	−0.02555	36.872322	0.01685
	7.745356	−0.15749	12.824138	0.05794
	2.681252	0.13375	4.460216	0.20687
	0.928184	0.61912	1.551256	0.39377
	0.321314	1.00000	0.539525	0.40850
			0.187646	0.18305

state at equilibrium geometry. A wide set of polarization functions were obtained [from 1d to (3d, 2f)]. The optimization of the d exponents were carried out with a (11s, 7p) basis set in an all-electron calculation and transferred to the (5s, 6p) basis set to be used along with the pseudopotential. The f exponents were optimized with the (11s, 7p, 3d) basis set in an all-electron calculation and were also transferred to the (5s, 6p, 3d) basis set. The exponents and the total energy of the homonuclear diatomic molecules are shown in Table III.

The diffuse functions (α_{n+1}) were defined in terms of the geometric progression based on the

TABLE III.
Exponents of Polarization Functions and Total SCF Energy of Diatomic Molecules Calculated with a (11s, 7p) Basis Function for d Exponents and with a (11s, 7p, 3d) for Calculations Including f Functions.

Molecule	Sets of Polarization Functions	Exponents	$-E_{\text{total}}$ (a.u.)
$\text{B}_2(^3\Sigma_g^-)$	1d	0.432519	49.0760672
		0.134269	49.0780430
		0.527598	
	3d	0.124495	49.0783318
		0.314569	
		0.672284	
	1f	0.427989	49.0793825
	2f	0.117571	49.0801437
$\text{C}_2(^1\Sigma_g^-)$	1d	0.715473	75.3848020
		0.248444	75.3871245
		0.886041	
	3d	0.201968	75.3875181
		0.597648	
		1.312612	
	1f	0.844879	75.3913584
	2f	0.201016	75.3921611
$\text{N}_2(^1\Sigma_g^-)$	1d	0.972478	108.9608139
		0.393172	108.9675786
		1.339417	
	3d	0.322941	108.9687820
		1.086958	
		3.642437	
	1f	1.319499	108.9720737
	2f	0.550841	108.9734357
$\text{O}_2(^3\Sigma_g^-)$	1d	0.942442	149.6471397
		0.244876	149.6543144
		1.155011	
	3d	0.176008	149.6571960
		0.775188	
		2.219958	
	1f	1.069305	149.6636713
	2f	0.254198	149.6656252
$\text{F}_2(^1\Sigma_g^-)$	1d	0.840153	198.7305398
		0.266068	198.734002
		1.078052	
	3d	0.199798	198.7353721
		0.789523	
		3.180636	
	1f	0.837999	198.7411688
	2f	0.377519	198.7423297
		1.415074	

two most diffuse functions of the original basis set²² as:

$$\alpha_{n+1} = \frac{\alpha_n^2}{\alpha_{n-1}} \quad (7)$$

where α_n is the smallest exponent of the basis set and α_{n-1} is the second smallest exponent.

The main difference between the all-electron (GCDZ) and SBK (GCDZ-SBK) double-zeta basis set lies in the core region. The basis set to be used along with the pseudopotential describes only the electronic distribution in the valence region, whereas the GCDZ basis set has an additional subset of function describing the core region.

The s and p exponents of GCDZ and GCDZ-SBK are not equal. However, this is not observed in the CEP-31G basis sets.^{12,13} In fact, the procedure to construct the GCDZ-SBK atomic basis sets reported in the present work is distinct from that employed in the literature. In refs. 12 and 13 the investigators used the same philosophy adopted by Pople and collaborators. In the N-31G basis sets, s- and p-shared exponent and contraction coefficients were obtained by atomic energy minimization. There is no indication that either the atomic orbitals or other atomic properties were applied to model the basis sets. In the development of the GCM-IOD basis sets, besides atomic energy properties, other one-electron molecular properties were also used along with the optimization process.¹⁹ This same philosophy has been applied for GCDZ-SBK basis sets, where two-electron molecular properties were also included for comparisons of theoretical and experimental results.

Results and Discussion

MOLECULAR CALCULATIONS AT THE HARTREE-FOCK LEVEL

The similarity in the development of N31G and CEP-31G suggests the use of the 6-31G* calculations as a reference to compare the performance of CEP-31G* basis sets to reproduce the all-electron molecular properties. The same comparison will be established between the GCM-IOD basis sets: (11s, 7p) \rightarrow [4s, 2p] for all-electron (GCDZ) and (5s, 6p) \rightarrow [2s, 2p] for SBK-ECP calculations (GCDZ-SBK). Five molecules containing first row elements were chosen to study the equivalence between all-electron and pseudopotential properties: C_2H_6 ; CH_3NH_2 ; CH_3OH ; CH_3F ; and BF_3 . Six

molecular properties were focused upon in the study: geometrical parameters (bond distances and bond angles); HOMO and LUMO energies; harmonic vibrational frequencies; IR intensities; and dipole moments. The comparison between all-electron and pseudopotential is not rigorously meaningful for dipole moments and infrared intensities, because the SBK pseudo-orbitals do not present the correct nodal structure and should be orthogonalized to the core. However, the comparison between the all-electron and pseudopotential results are presented indicating that the effect of orthogonalization does not significantly change these properties.

Four basis sets were analyzed at the Hartree–Fock level: DZ (double-zeta); DZ(1d); DZ + (1d); and DZ + (2d). The polarization and diffuse functions were employed only in the heavy atoms. The properties were calculated with optimized geometry up to 10^{-7} a.u. in the gradient for each of the basis sets.

The degree of equivalence between the pseudopotential and all-electron properties was calculated as a mean error (ME):

$$ME(\%) = 100 * \sum_i \frac{|AE_i - ECP_i|}{RF_i} \quad (8)$$

where AE_i and ECP_i are the values obtained with the all-electron basis sets (6-31G and GCDZ) and the pseudopotential basis sets (CEP-31G and GCDZ-SBK), respectively. RF_i is the reference value of the i th property obtained with the high-quality all-electron basis sets 6-311 + + G(3df, 3pd) and GCDZ-SBK + + (3df, 3pd). In addition to the comparison with the similar all-electron basis set, the pseudopotential calculations were also compared with high-level basis set results. We chose two high-quality basis sets to evaluate the molecular properties for this purpose, 6-311 + + G(3df, 3pd) and GCDZ-SBK + + (3df, 3pd), depending on the type of basis set employed (e.g., shared exponent or GCM-IOD).

Comparison of the pseudopotential results with high-level all-electron calculations was made employing another simple statistical criterion, the mean difference (MD):

$$MD = \frac{\sum_i |RF_i - ECP_i|}{n} \quad (9)$$

The Hartree–Fock results will not be presented and can be obtained by request. The analysis of the

Hartree–Fock results will be carried out in terms of ME and MD.

Geometrical Parameters

The results of bond distances for the four basis sets [DZ; DZ(1d); DZ + (1d); DZ + (2d)] along with shared exponent and GCM-IOD formulations were compared with the high-quality 6-311 + + G(3df, 3pd) and GCDZ + + (3df, 3pd) all-electron results. MD is 13.6 mÅ for the four basis sets used with CEP-31G and the 6-311 + + G(3df, 3pd) results, whereas, for GCDZ-SBK and GCDZ + + (3df, 3pd) calculations, the MD is 11.9 mÅ. For bonds involving only the heavy atoms calculated with CEP-31G and GCDZ-SBK, the MDs between pseudopotential and the respective high-level basis sets are 19.7 mÅ and 13.3 mÅ, respectively. Better performance with CEP-31G results is observed when the bonds with hydrogens are included, because their MD is 8.5 mÅ, whereas for GCDZ-SBK the value is 10.7 mÅ.

Even when comparing the results of the best pseudopotential basis sets (two polarization functions and one diffuse function) with the 6-311 + + G(3df, 3pd) results, one observes a better performance of GCDZ-SBK + (2d) basis sets. In this case, the MD is 6.7 mÅ and for CEP-31 + G(2d) the MD is 9.1 mÅ.

The performance of the GCDZ-SBK basis sets for bond angles is better than that of CEP-31G for reproducing the high-level all-electron results. A smaller MD value among all basis sets calculated with GCDZ-SBK is found: 0.42° . In terms of the 2d plus diffuse function pseudopotential basis sets, the MD value for GCDZ-SBK + (2d) decays to 0.12° , whereas for CEP-31 + G(2d) the value is 0.30° .

In general, the ME associated with geometrical parameters can be analyzed by means of Table IV. Table IV shows a systematic decrease of ME for the CEP-31G basis set results with inclusion of polarization and diffuse functions. An abrupt fall in ME of around 0.1% is observed for GCDZ-SBK + (2d). The ME values for GCDZ-SBK are smaller than CEP-31G in all cases, which indicates a greater similarity between all-electron and pseudopotential results for GCM-IOD basis sets.

Orbital Energies

Table IV shows the ME results for HOMO and LUMO energies. The GCDZ-SBK basis sets present better performance in reproducing the all-electron

TABLE IV.

Mean Error (ME) of Geometric Parameters (Geom), HOMO and LUMO Energies, Dipole Moments, (μ), Harmonic Frequencies (ν), and IR Intensities for Different Basis Sets.

	Geom	HOMO	LUMO	μ	ν	IR
CEP	0.58	0.71	49.60	9.36	1.68	85.5
GDZ	0.21	0.66	5.95	5.89	0.77	36.8
CEP(1d)	0.43	0.73	59.10	8.32	1.08	135.0
GDZ(1d)	0.22	0.60	8.43	5.38	0.93	16.3
CEP + (1d)	0.37	0.69	8.06	0.82	1.00	120.0
GDZ + (1d)	0.20	0.36	0.65	2.31	0.60	17.2
CEP + (2d)	0.32	0.65	7.03	0.94	0.98	88.4
GDZ + (2d)	0.09	0.37	0.53	1.11	0.29	12.6

energies. For HOMO energies, the GCM-IOD basis sets give better agreement between pseudopotential and all-electron calculations than the comparison between 6-31G and CEP-31G basis sets. A significant discrepancy is observed for the LUMO energies, where ME values for GCDZ are one order of magnitude lower than the values for CEP-31G basis sets.

The comparison of pseudopotential with high-level basis set results also shows the best performance of the GCDZ-SBK basis sets. The MD value between CEP-31 + G(2d) and 6-311 + + G(3df, 3pd) for HOMO energy is around 3.0 mE_h, whereas the equivalent GCM-IOD basis sets give an ME of 1.3 mE_h. The performance of GCDZ-SBK is even better for LUMO energy compared with CEP-31G and is in excellent agreement with GCDZ + + (3df, 3pd) and also with 6-311 + + G(3df, 3pd). It is worth noticing the importance of the diffuse functions in reducing the LUMO energies.

Dipole Moments

The inclusion of a pair of s and p diffuse functions and the second d-type polarization function shows a significant decrease in the ME values for both shared-exponent and GCM-IOD-derived basis sets (see Table IV). The CEP-31G basis set gives a result closer to GCDZ-SBK when the diffuse function is included, although the second polarization function diminishes for GCDZ-SBK. A lower dipole moment calculated with additional polarization and diffuse functions along with the pseudopotential is observed for both CEP-31G and GCDZ-SBK.

The MD value between CEP-31 + G(2d) and 6-31 + + G(3df, 3pd) is 0.046 Debye, which is very similar to the GCDZ value of 0.043 Debye. In other words, the CEP-31 + G(2d) and GCDZ-SBK + (2d)

basis sets reproduce the dipole moment values of the corresponding all-electron basis sets with the same degree of accuracy.

Harmonic Vibrational Frequencies

All GCDZ-SBK basis sets yielded unscaled frequency ME values under 1% with respect to the reference all-electron calculations (see Table IV). There is a tendency of the pseudopotential and all-electron basis sets to provide similar frequencies when polarization and diffuse functions are present. The second polarization function significantly decreases ME for GCDZ-SBK basis sets (about 0.3%). However, there is no significant difference in ME when the second d-type function is added to CEP-31G. Actually, the presence of only one polarization function is responsible for an almost constant value of ME of around 1% for shared-exponent basis sets.

The ME values between pseudopotential and high-level basis sets are shown in Table IV. Frequencies calculated with CEP-31 + G(2d) when compared with those calculated using 6-311 + + G(3df, 3pd) yielded an ME of 1.2%, which is just 0.2% above the result obtained for ME between CEP-31 + G(2d) and 6-31 + G(2d). In the case of GCDZ-SBK + (2d), the ME for GCDZ + + (3df, 3pd) is 0.9% higher than GCDZ + (2d) (\cong 1.2%). In this sense, the pseudopotential basis sets have the same degree of similarity for frequencies in comparison with high-level basis set results.

Infrared Intensities

Table IV shows the low degree of similarity between the pseudopotential and all-electron shared-exponent basis sets results compared with GCDZ-SBK and GCDZ results. It is well known

that IR intensity is one of the most difficult properties to calculate accurately.³¹ The double-polarized plus diffuse functions basis sets yield the best adjustment between pseudopotential and all-electron calculations for both shared-exponent and GCM-IOD. However, the ME value for GCDZ-SBK + (2d) is seven times lower than the respective value of CEP-31 + G(2d).

The ME values presented in Table IV for IR intensities show that GCDZ-SBK basis sets reproduce high-level results with better agreement than the CEP-31G basis sets. For both pseudopotential basis sets, the inclusion of polarization and diffuse functions diminish the differences with respect to the high-level basis set results. However, the ME value between GCDZ-SBK + (2d) and GCDZ++(3df,3pd) is two times lower than the respective value of CEP-31 + G(2d).

MOLECULAR CALCULATION AT MP2 LEVEL

The molecular properties were calculated using double-zeta basis sets plus two polarization and one diffuse function for both all-electron and pseudopotential approximations. The comparison between results is made and the degree of fitness is discussed in terms of eq. (8) and eq. (9). In this case, the available experimental results for the three properties, geometrical parameters, dipole moments, and fundamental vibrational frequencies, are employed as reference.

Geometrical Parameters

The ME values for bond distances of all molecules are lower when GCDZ-SBK and GCDZ results are analyzed. The mean value of ME for all molecules gives the same tendency for bond angles: for instance, the GCM-IOD-derived basis sets have a higher degree of similarity than the shared-exponent basis sets. ME comparisons for all molecules are summarized in Figure 2. Whereas, ME for GCDZ-SBK + (2d) is 0.35%, CEP-31 + G(2d) provides an ME of 0.88%, which is 2.4 times larger.

Table V presents the experimental³² and pseudopotential results for bond distances and bond angles. The GCDZ-SBK + (2d) basis set yielded more reliable results than the CEP-31 + G(2d) basis set. The MD for bond distances calculated with GCDZ-SBK + (2d) was 9.0 mÅ, whereas for CEP-31 + G(2d) was 13.4 mÅ. The C—C bond is a particular case of the reliability of GCDZ-SBK + (2d) in reproducing experimental data. Whereas

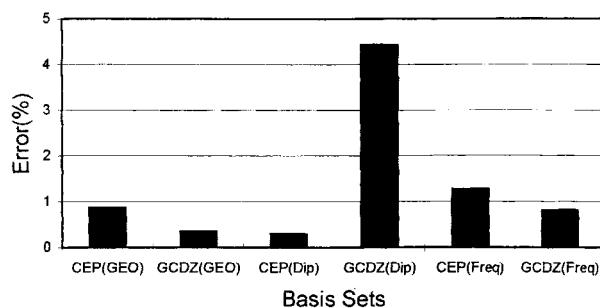


FIGURE 2. Mean errors of geometrical parameters (GEO), dipole moments (Dip), and frequencies (Freq) for CEP-31 + G(2D) (CEP) and GCDZ-SBK + (2d) (GCDZ) at the MP2 level of calculation.

the GCDZ-SBK + (2d) basis set showed a difference of 2.9 mÅ, with respect to experimental results, the CEP-31 + G(2d) basis set provided a difference of 21.2 mÅ. The O—H experimental bond distance is not well reproduced by either basis set, although the theoretical results are close to each other.

The GCDZ-SBK + (2d) basis set again yielded bond angles in close agreement with experimental data. The shared-exponent basis set for the pseudopotential provided an MD of 0.41° for all molecules, whereas GCDZ-SBK presented 0.24°. In CH₃F, the HCF angle was changed to HCH to compare with the experimental results. The ME values for both basis sets were very similar: 0.24% for CEP-31 + G(2d), and 0.18% for GCDZ-SBK + (2d).

Dipole Moments

Experimental³² and theoretical results for dipole moments are given in Table VI. For CH₃OH and CH₃F, the GCDZ-SBK + (2d) provides dipole moment values in closer agreement with experiment, whereas CEP-31 + G(2d) maintains a difference from experimental results by roughly 0.2 Debye. The worst agreement with experimental data was observed for CH₃NH₂, where both CEP-31 + G(2d) and GCDZ-SBK + (2d) overestimated the result by around 0.2 Debye. However, there was no significant difference between the theoretical values for this molecule. Thus, GCDZ-SBK + (2d) yielded an MD value of 0.08 Debye, which suggests better reproduction of the experimental results compared with CEP-31 + G(2d), which yielded 0.17 Debye for MD.

The ME analysis (Fig. 2) provided a different tendency for this property. In this case, the expo-

TABLE V.
Geometrical Parameters (Angstroms and Degrees) for Some Molecules Obtained at MP2 Level of Theory.

Molecule	Parameter	6-31G + (2d)	GCDZ + (2d)	CEP-31G + (2d)	GCDZ-SBK + (2d)	Exp. (ref. 30)
C ₂ H ₆	<i>r</i> _{CC}	1.5307	1.5256	1.5492	1.5309	1.528
	<i>r</i> _{CH}	1.0971	1.1038	1.1042	1.1014	1.088
	<i>A</i> _{HCC}	111.46	111.31	111.22	111.16	111.6
CH ₃ NH ₂	<i>r</i> _{CN}	1.4669	1.4568	1.4797	1.4670	1.471
	<i>r</i> _{CH}	1.0979	1.1062	1.1052	1.1030	1.099
	<i>r</i> _{NH}	1.0155	1.0211	1.0224	1.0235	1.010
	<i>A</i> _{HNC}	110.92	110.08	110.60	110.50	110.3
CH ₃ OH	<i>r</i> _{CO}	1.4302	1.4240	1.4386	1.4213	1.425
	<i>r</i> _{CH}	1.0968	1.1001	1.1044	1.1022	1.094
	<i>r</i> _{OH}	0.9667	0.9719	0.9746	0.9739	0.945
	<i>A</i> _{HOC}	107.83	108.06	107.42	108.24	108.3
CH ₃ F	<i>r</i> _{CF}	1.4052	1.3925	1.4015	1.3915	1.383
	<i>r</i> _{CH}	1.0936	1.0973	1.1015	1.0985	1.100
	<i>A</i> _{HCH}	110.83	110.36	110.55	110.36	110.6
BF ₃	<i>r</i> _{BF}	1.3215	1.3219	1.3221	1.3212	1.310

nent basis set for the pseudopotential and all-electron basis sets presented greater similarity. The ME value for these basis sets was 0.3%, whereas for GCDZ it was 4.4%. A great discrepancy was again observed for CH₃NH₂, where GCDZ + (2d) overestimated the experimental result by 0.13 Debye (10%).

Harmonic Vibrational Frequencies

Analysis of unscaled vibrational frequencies requires comparison of a wide range of numbers. In this sense, comparison between theoretical and experimental data should be made in terms of ME, where a percentage value is obtained.

Table VII shows ME for each molecule and the respective average value with respect to the harmonic experimental fundamental vibrational values³⁴. For each molecule, the GCDZ-SBK + (2d)

basis set provided results in closer agreement with experimental measurements than the CEP-31 + G(2d) basis set. The average ME difference between GCDZ-SBK + (2d) and CEP-31 + G(2d) was 1.5%. To correct the theoretical pseudopotential result to reproduce the experimental data, one could employ a value of 97.4% for GCDZ-SBK + (2d) and 95.9% for CEP-31 + G(2d) results. The use of ME analysis to compare pseudopotential and all-electron results again shows that the best performance was achieved with GCM-IOD-derived basis sets (see Fig. 2).

Conclusion

A new approach for the development of basis sets to be used along with effective core potentials was systematically studied. The main criterion was

TABLE VI.
Dipole Moments (Debye) for Some Molecules Calculated at MP2 Level of Theory.

Molecule	6-31G + (2d)	GCDZ + (2d)	CEP-31G + (2d)	GCDZ-SBK + (2d)	Exp. (ref. 33)
CH ₃ NH ₂	1.483	1.623	1.477	1.491	1.29
CH ₃ OH	1.893	1.762	1.894	1.715	1.71
CH ₃ F	2.005	1.884	1.996	1.877	1.85

TABLE VII.
Mean Error (%) for Harmonic Vibrational Frequencies
for Ground State of Some Polyatomic Molecules
Calculated with Different Basis Sets at MP2 Level
of Theory.

Molecule	ME CEP	ME GCDZ
C ₂ H ₆	0.92	0.96
CH ₃ NH ₂	0.68	1.49
CH ₃ OH	1.05	0.57
CH ₃ F	0.71	0.28
BF ₃	2.97	0.79
Mean	1.27	0.82

Experimental results obtained from ref. 34.

the observation of the tendency of the LCAO coefficients versus the $\ln(\alpha)$ of the respective primitives. This picture can be associated with a model called generator coordinate and can provide a visual criterion to define the range in which the basis set should be developed or modified. Double-zeta basis sets were modeled for an SBK pseudopotential from all-electron basis sets for a set of compounds containing elements of the second period of the periodic table. Application of the modeled basis sets at the Hartree-Fock level of theory for SBK and all-electron calculations revealed an excellent adjustment between both sets of calculations for different properties such as geometries, HOMO and LUMO energies, dipole moments, and vibrational frequencies. At the MP2 level, the results obtained with the modeled basis sets were also in good agreement with experimental data. The results analyzed showed that the new method is reliable and can provide a simple and useful alternative for development of basis sets. Applications of the GCDZ-SBK basis sets in larger molecular systems could provide reliable results, compared with all-electron and experimental measurements, and could also help minimize computational costs.

Acknowledgments

We are thankful for the use of the computational facilities at Cenapad (Centro Nacional de Processamento de Alto Desempenho), São Paulo, where the Gaussian/94³⁵ and Gamess/95³⁶ programs were conveniently adapted and used for the calculations.

References

1. P. Neogrády, V. Kellö, M. Urban, and A. J. Sadlej, *Theor. Chim. Acta*, **93**, 101 (1996).
2. R. E. Easton, D. J. Giesen, A. Welch, C. J. Cramer, and D. G. Truhlar, *Theor. Chim. Acta*, **93**, 28 (1996).
3. H. Tatewaki and A. J. Thakkar, *J. Chem. Phys.*, **104**, 8493 (1996).
4. T. Koga, H. Tatewaki, and A. J. Thakkar, *J. Chem. Phys.*, **100**, 8140 (1994).
5. R. A. Kendall and T. H. Dunning, Jr., *J. Chem. Phys.*, **96**, 6796 (1992).
6. C. F. Bunge, J. A. Barrientos, A. V. Bunge, and J. A. Cogordan, *Atom. Data Nucl. Data Tables*, **53**, 118 (1993).
7. M. Dolg, *Theor. Chim. Acta*, **93**, 141 (1996).
8. P. Schwerdtfeger, T. Fischer, M. Dolg, G. Igel-Mann, A. Nicklass, H. Stoll, and A. Haaland, *J. Chem. Phys.*, **102**, 2050 (1995).
9. T. Leininger, A. Nicklass, H. Stoll, M. Dolg, and P. Schwerdtfeger, *J. Chem. Phys.*, **105**, 1052 (1996).
10. P. J. Hay and W. R. Wadt, *J. Chem. Phys.*, **82**, 270, 299 (1985).
11. W. R. Wadt and P. J. Hay, *J. Chem. Phys.*, **82**, 284 (1985).
12. W. J. Stevens, H. Basch, and M. Krauss, *J. Chem. Phys.*, **81**, 6026 (1984).
13. W. J. Stevens, M. Krauss, H. Basch, and P. G. Jasien, *Can. J. Chem.*, **70**, 612 (1992).
14. D. L. Hill and J. A. Wheeler, *Phys. Rev.*, **89**, 1102 (1953).
15. J. J. Griffin and J. A. Wheeler, *Phys. Rev.*, **108**, 311 (1957).
16. J. R. Mohallem, R. M. Dreizler, and M. Trsci, *Int. J. Quantum Chem. Symp.*, **20**, 45 (1986).
17. H. F. M. da Costa, A. M. Simas, V. H. Smith, Jr., and M. Trsic, *Chem. Phys. Lett.*, **192**, 195 (1992).
18. R. Custodio, M. Giordan, N. H. Morgon, and J. D. Goddard, *Int. J. Quantum Chem.*, **42**, 411 (1992).
19. R. Custodio, J. D. Goddard, M. Giordan, and N. H. Morgon, *Can. J. Chem.*, **70**, 580 (1992).
20. R. Custodio and J. D. Goddard, *J. Mol. Struct. (THEOCHEM)*, **277**, 263 (1992).
21. R. O. Vianna, R. Custodio, H. Chacham, and J. R. Mohallen, *Int. J. Quant. Chem. Symp.*, **26**, 311 (1992).
22. R. Custodio and J. D. Goddard, *J. Mol. Struct. (THEOCHEM)*, **281**, 75 (1993).
23. R. Custodio, W. M. Davis, and J. D. Goddard, *J. Mol. Struct. (THEOCHEM)*, **315**, 163 (1994).
24. N. H. Morgon, R. Custodio, and J. M. Riveros, *Chem. Phys. Lett.*, **235**, 436 (1995).
25. N. H. Morgon, *J. Phys. Chem.*, **99**, 17832 (1995).
26. J. C. Phillips and L. Kleinman, *Phys. Rev.*, **16**, 287 (1959).
27. V. Bonifacic and S. Huzinaga, *J. Chem. Phys.*, **60**, 2779 (1974).
28. A. F. R. de Toledo Piza, E. J. V. de Passos, D. Galetti, M. C. Nemes, and M. M. Watanabe, *Phys. Rev. C*, **15**, 1477 (1977).
29. M. Giordan, *Master Thesis*, Instituto de Química, Universidade Estadual de Campinas, Campinas, 1992.
30. T. H. Dunning, *J. Chem. Phys.*, **53**, 2823 (1970).
31. A. L. M. S. de Azevedo, B. B. Neto, I. S. Scarmino, A. E. de Oliveira, and R. E. Bruns, *J. Comput. Chem.*, **17**, 167 (1996).

32. C_2H_6 : J. L. Duncan, D. C. McKean, and J. Bruce, *J. Molec. Spectrosc.*, **74**, 361 (1979); CH_3NH_2 : K. Takagi and T. Kojima, *J. Phys. Soc. Jpn.*, **30**, 1145 (1971); CH_3OH : R. M. Lees and J. C. Backer, *J. Chem. Phys.*, **48**, 5299 (1968); CH_3F : W. W. Clark and F. C. deLucia, *J. Mol. Struct.*, **32**, 29 (1976); BF_3 : S. G. W. Ginn, J. K. Kenny, and J. Overend, *J. Chem. Phys.*, **48**, 1571 (1968).
33. A. L. McClellan, *Tables of Experimental Dipole Moments*, W. H. Freeman, San Francisco, 1963.
34. C_2H_6 : F. Stitt, *J. Chem. Phys.*, **7**, 297 (1939); CH_3NH_2 : T. Shimanouchi, *Tables of Molecular Vibrational Frequencies, Vol. 1*, NSRDS-NBS 39 (1972); CH_3OH : A. Trinider and G. Zerbi, *Zeit. Naturforschung*, **A25**, 1729 (1970); CH_3F : T. Shimanouchi, *Tables of Molecular Vibrational Frequencies, Vol. 1*, NSRDS-NBS 39 (1972); BF_3 : T. Shimanouchi, *J. Phys. Chem. Ref. Data*, **6**, 993 (1972).
35. M. J. Frisch, G. W. Trucks, H. B. Schlegel, P. M. W. Gill, B. G. Johnson, M. A. Robb, J. R. Cheesemann, T. Keith, G. A. Petersson, J. A. Montgomery, K. Rahavachari, M. A. Al-Laham, V. G. Zakrzewski, J. V. Ortiz, J. B. Foresman, J. Cioslowski, B. B. Stefanov, A. Nanayakkara, M. Challacombe, C. Y. Peng, P. Y. Ayala, W. Chen, M. W. Wong, J. L. Andres, E. S. Replogle, R. Gomperts, R. L. Martin, D. J. Fox, J. S. Binkley, D. J. Defrees, J. Baker, J. P. Stewart, M. Head-Gordon, C. Gonzalez, and J. A. Pople, Gaussian-94, Rev. D.2, Gaussian, Inc., Pittsburgh, PA, 1995.
36. Gamess 95, M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Kosek, N. Matsunaga, K. A. Nguyen, S. J. Su, T. L. Windus, M. Dupuis, and J. A. Montgomery, *J. Comput. Chem.*, **14**, 1347 (1993).